WHAT IS CLAIMED IS:

1. A system for temporal modification of segments of an audio signal, comprising:

extracting data frames from an audio signal;

examining content of each data frame and classifying a type of each data frame according to pre-established criteria;

temporally modifying at least part of at least one of the data frames using a temporal modification process that is specific to the classification type of each data frame.

2. The system of claim 1 wherein the classification of frame type is based solely on the frame being classified.

3. The system of claim 1 wherein the classification of frame type is at least partially based on information derived from one or more neighboring frames.

4. The system of claim 1 wherein the frames are processed sequentially.

5. The system of claim 1 wherein the classification is at least partially based on a periodicity of each data frame.

6. The system of claim 1 wherein the frame types include voiced frames and unvoiced frames.

7. The system of claim 6 wherein the frame types further include mixed frames, said mixed frames including both voiced and unvoiced segments.

8. A method for temporal modification of segments of an audio signal including speech, comprising:

sequentially extracting data frames from a received audio signal;

determining a content type of each segment of a current frame of the sequentially extracted data frames, said content types including voiced segments, unvoiced segments, and mixed segments;

temporally modifying at least one segment of the current frame by automatically selecting and applying a corresponding temporal modification process for the at least one segment of the current frame from among a voiced segment temporal modification process, an unvoiced temporal modification process, and a mixed segment temporal modification process.

9. The method of claim 8 further comprising estimating an average pitch period for each frame, said frames each comprising at least one segment of approximately one pitch period in length;

10. The method of claim 8 wherein determining the content type of each segment of the current frame comprises computing a normalized cross correlation for each frame and comparing a maximum peak of each normalized cross correlation to predetermined thresholds for determining the content type of each segment.

11. The method of claim 8 wherein the content type of at least one segment is a voiced segment, and wherein temporally modifying the at least one segment comprises stretching the voiced segment to increase a length of the current frame.

12. The method of claim 11 wherein stretching the voiced segment comprises:

identifying at least one of the segments as a template;

searching for a matching segment whose cross correlation peak exceeds a predetermined threshold; and

aligning and merging the matching segments of the frame.

13. The method of claim 12 wherein identifying at least one of the segments as a template comprises selecting a template from the end of the frame, and wherein searching for the matching segment comprises examining a recent past of the audio signal to identify a match.

14. The method of claim 12 wherein identifying at least one of the segments as a template comprises selecting a template from the beginning of the frame, and wherein searching for the matching segment comprises examining a near future of the audio signal to identify a match.

15. The method of claim 12 wherein identifying at least one of the segments as a template comprises selecting a template from between the beginning and end of the frame, and wherein searching for the matching segment comprises examining a near future and a near past of the audio signal to identify a match.

16. The method of claim 12 further comprising alternating selection points for the template such that consecutive templates are identified at different positions within the current frame.

17. The method of claim 8 further comprising determining whether an average compression ratio of temporally modified segments corresponds to an overall target compression ratio, and wherein a next target compression ratio for at least one next current frame is automatically adjusted as needed for ensuring that the overall target compression ratio is approximately maintained.

18. The method of claim 8 wherein the content type of at least one segment is an unvoiced segment, and wherein temporally modifying the at least one segment comprises automatically generating and inserting at least one synthetic segment into the current frame to increase a length of the current frame.

19. The method of claim 18 wherein automatically generating the at least one synthetic segment comprises automatically computing the Fourier transform the current frame, introducing a random rotation of the phase into the FFT coefficients, and then computing the inverse FFT for each segment, thereby creating the at least one synthetic segment.

20. The method of claim 8 wherein the content type of at least one segment is a mixed segment, and wherein the mixed segment includes both voiced and unvoiced components.

21. The method of claim 20 wherein temporally modifying the mixed segment comprises:

identifying at least one of the segments as a template;

searching for a matching segment whose cross correlation peak exceeds a predetermined threshold;

aligning and merging the matching segments of the frame to create an interim voiced segment;

automatically generating and inserting at least one synthetic segment into the current frame to create an interim unvoiced segment;

weighting each of the interim voiced segment and the interim unvoiced segment relative to a normalized cross correlation peak computed for the current segment; and

adding and windowing the interim voiced segment and the interim unvoiced segment to create a partially synthetic stretched segment.

22. The method of claim 8 wherein the content type of at least one segment is a voiced segment, and wherein temporally modifying the at least one segment comprises compressing the voiced segment to decrease a length of the current frame.

23. The method of claim 22 wherein compressing the voiced segment comprises:

identifying at least one of the segments as a template;

searching for a matching segment whose cross correlation peak exceeds a predetermined threshold;

cutting out the signal between the template and the match; and

aligning and merging the matching segments of the frame.

24. The method of claim 8 wherein the content type of at least one segment is an unvoiced segment, and wherein temporally modifying the at least one segment comprises compressing the unvoiced segment to decrease a length of the current frame.

25. The method of claim 24 wherein compressing the voiced segment comprises:

shifting a segment of the frame from a first position in the frame to a second position in the frame;

deleting the portion of the frame between the first position and the second position; and

adding the shifted segment of the frame to the signal representing the remainder of the frame by using a sine windowing function for blending the edges of the segment with the signal representing the remainder of the frame.

26. A computer-implemented process for providing dynamic temporal modification of segments of a digital audio signal, comprising using a computing device to:

receive one or more sequential frames of a digital audio signal;

decode each frame of the digital audio signal as it is received;

determine a content type of segments of the decoded audio signal from a group of predefined segment content types, each segment content type having an associated type-specific temporal modification process; and

modify a temporal scale of one or more segments of the decoded audio signal using the associated type-specific temporal modification process specific to each segment content type.

27. The computer-implemented process of claim 26 wherein the group of predefined segment content types includes voiced type segments and unvoiced type segments.

28. The computer-implemented process of claim 27 wherein the group of predefined segment content types further includes mixed type segments, said mixed type segments representing a mixture of voiced content and unvoiced content.

29. The computer-implemented process of claim 27 wherein modifying the temporal scale of one or more segments comprises any of temporally stretching and temporally compressing the one or more segments to approximately achieve a target temporal modification ratio.

30. The computer-implemented process of claim 29 wherein the target temporal modification ratio of subsequent segments is automatically adjusted to achieve an average target temporal modification ratio relative to actual temporal scale modification of at least one preceding segment.

31. The computer-implemented process of claim 27 wherein determining the content type of segments comprises computing a normalized cross correlation for sub-segments of each segment, and comparing a maximum

peak of each normalized cross correlation to predetermined thresholds for determining the content type of each segment.

32. The computer-implemented process of claim 27 wherein at least one segment is a voiced type segment, and wherein modifying the temporal scale of voiced type segments comprises stretching at least one voiced type segment by approximately one or more pitch periods to increase a length of the at least one voiced type segment.

33. The computer-implemented process of claim 27 wherein stretching the at least one voiced type segment comprises:

identifying at least one sub-segment of approximately one pitch period in length as a template;

searching for a matching sub-segment whose cross correlation peak exceeds a predetermined threshold; and

aligning and merging the matching segments of the frame.

34. The computer-implemented process of claim 27 wherein at least one segment is an unvoiced type segment, and wherein modifying the temporal scale of unvoiced type segments comprises:

automatically generating at least one synthetic segment from one or more sub-segments of the at least one unvoiced-type segment; and

inserting the at least one synthetic segment into the at least one unvoiced type segment to increase a length of the at least one unvoiced type segment.

35. The computer-implemented process of claim 34 wherein automatically generating the at least one synthetic segment comprises:

automatically computing the Fourier transform of at least one sub-segment of the at least one unvoiced type segment;

randomizing the phase of at least some of the computed FFT coefficients; and

computing the inverse FFT for the computed FFT coefficients to generate the at least one synthetic segment.

36. The computer-implemented process of claim 34 further comprising automatically determining one or more insertion points for inserting the at least one synthetic segment into the at least one unvoiced type segment.